

NON-PROVISIONAL
PATENT APPLICATION

***FUSOBACTERIUM* NUCLEIC ACIDS, PLASMIDS AND VECTORS**

Inventor(s):

Susan A. Kinder Haake, a citizen of the United States of America, residing at
5289 Dobson Way
Culver City, CA 90230
United States of America

Kara K. Podkaminer, a citizen of the United States of America, residing at
Kinderlane
Jamesville, NY 13078
United States of America

Gwynne Attarian, a citizen of the United States of America, residing at
1133 Nielson Ct. #5
Ann Arbor, MI 48105
United States of America

Sean C. Yoder, a citizen of the United States of America, residing at
11454 North Poema Place, #103
Chatsworth, CA 91311
United States of America

Assignee:

Regents of University of California
Office of Technology Transfer
1111 Franklin Street, 5th Floor
Oakland, CA 94607-5200

FUSOBACTERIUM NUCLEIC ACIDS, PLASMIDS AND VECTORS

CROSS-REFERENCES TO RELATED APPLICATIONS

This application claims priority to provisional application U.S.S.N. 60/173,168, filed December 27, ~~2000~~ 1999, the disclosure of which is herein incorporated by reference in its entirety.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

This invention was made with Government support under Grant Nos. DE11180 and DE12639, awarded by the National Institutes of Health. The Government has certain rights in this invention.

FIELD OF THE INVENTION

The invention provides origin of replication sequences and replication genes and proteins for a plasmid functional in *Fusobacterium* (e.g., *F. nucleatum*). Provided by the invention are also plasmids and vectors that can replicate in *Fusobacterium* and related species. Further, the invention provides shuttle vectors that can replicate in *Fusobacterium* and in other microorganisms, such as *E. coli*. Still further, the present invention provides host cells comprising the plasmids and shuttle vectors, and methods for transformation of the host cells with the plasmid and shuttle vectors of the invention.

BACKGROUND OF THE INVENTION

Fusobacterium species are anaerobic Gram-negative microorganisms that are commonly found in the mouth. *Fusobacterium nucleatum* is the most frequently isolated pathogen from periodontal disease sites and is believed to be an initiator of periodontal diseases (see, e.g., Moore & Moore, *Periodontology* 2000 5:66-77 (1994)). Moreover, this bacterium is commonly found in abscesses and other infections in the abdomen, blood, chest, lung, sinuses, and female genital tract (see, e.g., Brook, *J. Clin. Microbiol.* 26:1181-1188 (1988); Holst et al., *J. Clin. Microbiol.* 32:176-186 (1994); Moore & Moore, *supra*). *F. nucleatum* is usually found as a component of polymicrobial infections, but is also found as a single isolate from infections, demonstrating its pathogenic potential. The virulence properties of *F. nucleatum* are related to its adherence to host tissues and other bacterial species, as well as its modulation of host cell

immune function (see, Bolstad, *Clin. Microb.* 9:55-71 (1996)) Thus, *F. nucleatum* is a microorganism of interest for further investigation because of its role in the periodontal diseases and other human infectious diseases.

Thus far, several *F. nucleatum* proteins that are believed to be associated with *F. nucleatum* pathogenesis have been cloned and expressed in *Eschericia coli*. For example, a *fomA* gene that encodes the major outer membrane protein has been cloned. This protein functions as a porin and may act as a receptor in coaggregation with other pathogenic bacteria (see, Jensen *et al.*, *Microb. Path.* 21:331-342 (1996); Kinder Haake *et al.*, *Arch. Oral Biol.* 42:19-24 (1997)). Also cloned is a *fipA* gene which encodes an immunosuppressive factor that inhibits human T-cell-responses (see, Demuth *et al.*, *Infect. Immun.* 64:1335-1341 (1996)). Moreover, a homologous family of small cryptic plasmids in strains of *F. nucleatum* has been isolated (McKay *et al.*, *Plasmid* 33:15-20 (1995)). However, the molecular manipulation of genes in *F. nucleatum* has not been possible due to the lack of reliable gene transfer and expression systems for *F. nucleatum*. The development of gene transfer and expression systems for *F. nucleatum* is essential to express and fully characterize the cloned *F. nucleatum* genes in the native host background.

Therefore, there is a need for reliable gene transfer and expression systems for *F. nucleatum*. The availability of gene transfer and expression systems would assist the molecular characterization of genes that are associated with pathogenesis of *F. nucleatum*. The availability of such systems would further assist the development of compounds that can prevent or treat periodontal diseases or other human infectious diseases caused by *F. nucleatum*. The present invention fulfills this and other needs.

SUMMARY OF THE INVENTION

The present invention provides for the first time origin of replication sequences, repA nucleic acids, and RepA polypeptides of a plasmid functional in *Fusobacterium*, in particular *F. nucleatum*. Also provided by the invention are plasmids and vectors that can replicate in *Fusobacterium*. Further, the invention provides shuttle vectors that can replicate in *Fusobacterium* and in other microorganisms, such as *E. coli*. Still further, the present invention provides host cells comprising these plasmids and shuttle vectors, and methods for transformation of the host cells. Embodiments of the present invention would assist cloning of *F. nucleatum* genes as well as the molecular characterization of genes that are associated with pathogenesis of *F. nucleatum*.

Moreover, embodiments of the invention can be used to express foreign genes, such as antigenic determinants of pathogenic microorganism, in *F. nucleatum*. Such transformants can be used as vaccine delivery systems to stimulate mucosal immunity against these pathogens.

5 In one aspect, the present invention provides isolated origin of replication sequences for *F. nucleatum* that comprise at least two copies of an iteron, wherein the iteron has a nucleic acid sequence of SEQ ID NO:3. In one embodiment, the isolated origin of replication sequences comprise two to six copies of the iteron having a sequence of SEQ ID NO:3. In another embodiment, the isolated origin of replication sequences
10 comprise a nucleic acid sequence of SEQ ID NO:4. In yet another embodiment, the isolated origin of replication sequences comprise a nucleic acid sequence of nucleotide position 3936 to 4481 of plasmid pFN1.

In another aspect, the present invention provides an isolated repA nucleic acid for *F. nucleatum*, wherein the repA nucleic acid encodes a protein that has greater
15 than about 80% amino acid sequence identity to a polypeptide having a sequence of SEQ ID NO:1, or to a polypeptide that selectively binds to polyclonal antibodies generated against SEQ ID NO:1. In one embodiment, the isolated repA nucleic acid encodes a polypeptide having a sequence of SEQ ID NO:1. In another embodiment, the isolated repA nucleic acid encodes a polypeptide that has a molecular weight of about 44.8 kDa.
20 In yet another embodiment, the isolated repA nucleic acid has a sequence of SEQ ID NO:2.

In yet another aspect, the present invention provides an isolated nucleic acid molecule comprising a DNA fragments of plasmid pFN1 or plasmid pFN2. In one embodiment, the isolated nucleic acid molecule comprises a 2.36 kb DNA fragment
25 generated by cleaving plasmid pFN1 with restriction endonucleases *AvrII* and *ScaII*. In another embodiment, the isolated nucleic acid molecule comprises a 0.9 kb DNA fragment generated by cleaving plasmid pFN2 with restriction endonucleases *HincII* and *HpaII*.

In yet another aspect, the present invention provides an isolated RepA
30 protein that has greater than about 80% amino acid sequence identity to a polypeptide having a sequence of SEQ ID NO:1, or to a polypeptide that selectively binds to polyclonal antibodies generated against SEQ ID NO:1. In one embodiment, the isolated RepA protein has greater than about 97% amino acid sequence identity to a polypeptide

having an amino acid sequence of SEQ ID NO:1. In another embodiment, the isolated RepA protein has the amino acid sequence of SEQ ID NO:1.

In yet another aspect, the present invention provides an isolated or recombinant plasmid comprising an origin of replication that comprises at least two
5 copies of an iteron having a nucleic acid sequence of SEQ ID NO:3. In one embodiment, the plasmid comprises an origin of replication that comprises between two to six copies of the iteron having a nucleic acid sequence of SEQ ID NO:3. In another embodiment, the plasmid comprises an origin of replication that comprise a nucleic acid sequence of SEQ ID NO:4.

10 In yet another aspect, the present invention provides an isolated plasmid comprising a repA nucleic acid, wherein the repA nucleic acid encodes a protein that has greater than about 80% amino acid sequence identity to a polypeptide having a sequence of SEQ ID NO:1, or to a polypeptide that selectively binds to polyclonal antibodies generated against SEQ ID NO:1, provided that the nucleic acid encoding the RepA
15 protein has other than the nucleic acid sequence of SEQ ID NO:5. In one embodiment, the plasmid comprises a repA nucleic acid that encodes a polypeptide having a sequence of SEQ ID NO:1. In another embodiment, the plasmid comprises a repA nucleic acid that has a sequence of SEQ ID NO:2. In yet another embodiment, the plasmid comprises a marker gene, preferably an antibiotic resistance gene.

20 In yet another aspect, the present invention provides an isolated or recombinant plasmid comprising any combination of origin of replication sequences and repA nucleic acids described herein. For example, in one embodiment, the plasmid comprises an origin of replication sequence comprising SEQ ID NO:4 and a repA nucleic acid that encodes a polypeptide having a sequence of SEQ ID NO:1. In another
25 embodiment, the plasmid further comprises a marker gene, preferably an antibiotic resistance gene.

In yet another aspect, the present invention provides an isolated or recombinant plasmid comprising a DNA fragment derived from plasmid pFN1 or plasmid pFN2. In one embodiment, the plasmid comprises a nucleic acid sequence of nucleotide
30 position 3936 to 4481 of plasmid pFN1. In another embodiment, the plasmid comprises a 2.36 kb DNA fragment which can be generated by cleaving plasmid pFN1 with restriction endonucleases *AvrII* and *ScaII*. In yet another embodiment, the plasmid comprises a 0.9 kb DNA fragment which can be generated by cleaving plasmid pFN2 with restriction endonucleases *AvrII* and *ScaII*.

In yet another aspect, the present invention provides an isolated plasmid designated pFN1 that has a GenBank Accession No. AF159249. This plasmid has partial restriction maps as shown in Figures 1A, 2A, 3 and 5.

5 In yet another aspect, the present invention provides an isolated plasmid designated pFN2 that have partial restriction maps as shown in Figures 1A, 3 and 5.

In yet another aspect, the present invention provides an isolated plasmid designated pFN3 that has a partial restriction map as shown in Figure 1A.

10 In yet another aspect, the present invention provides a shuttle vector comprising origin of replication sequences so that the vector can replicate in more than one microorganism. In one embodiment, the shuttle vector comprises an origin of replication functional in *E. coli* and an origin of replication functional in *F. nucleatum*, wherein the origin of replication functional in *F. nucleatum* comprises at least two copies of an iteron having a nucleic acid sequence of SEQ ID NO:3. In another embodiment, the shuttle vector comprises an origin of replication that comprises between two to six copies
15 of the iteron having a sequence of SEQ ID NO:3. In another embodiment, the shuttle vector comprises an origin of replication that comprises a sequence of SEQ ID NO:4. In yet another embodiment, the shuttle vector comprises an origin of replication that comprises a nucleic acid sequence of nucleotide position 3936 to 4481 of plasmid pFN1. In yet another embodiment, the shuttle vector further comprises a repA nucleic acid that
20 encodes a protein that has greater than about 80% amino acid sequence identity to a polypeptide having a sequence of SEQ ID NO:1, or to a polypeptide that selectively binds to polyclonal antibodies generated against SEQ ID NO:1. In yet another embodiment, the shuttle vector comprises a repA nucleic acid that encodes a protein having a sequence of SEQ ID NO:1. In yet another embodiment, the shuttle vector comprises a repA nucleic
25 acid that has a sequence of SEQ ID NO:2. In yet another embodiment, the shuttle vector further comprises a marker gene, preferably an antibiotic resistance gene, more preferably an *ermF-ermAM* cassette. In yet another embodiment, the shuttle vector further comprises a transcription cassette comprising a nucleic acid of interest operably linked to a promoter. In yet another embodiment, the shuttle vector is plasmid pHS17 that has a
30 nucleotide sequence of SEQ ID NO:15.

In yet another aspect, the invention provides host cells and methods for transformation of host cells with any one or more of plasmids and vectors described herein. In one embodiment, the host cell is *E. coli*. In another embodiment, the host cell is *F. nucleatum*.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A show partial restriction maps of *F. nucleatum* plasmids pFN1, pFN2, pFN3, and shuttle plasmid pHS17. Selected restriction endonuclease sites in the native plasmids are presented. Restriction endonuclease sites indicated for pHS17 relate to the plasmid construction. The pFN1 portion of pHS17 is indicated by the thick solid bar with the position of the *repA* homologue (ORF5) and putative *ori* indicated.

Figure 1B and Figure 1C show Southern blots of *F. nucleatum* plasmids. Plasmids from *F. nucleatum* strains 12230 (pFN1, lanes 1), 10113 (pFN2, lanes 2) and ATCC 10953 (pFN3, lanes 3) were probed with pFN1 (panel B; *EcoRI* digests) or pFN3 (panel C; *EcoRV* digests) DNA. The *HincII*-digested pFN1 and *AseI*-digested pFN3 probes were ³²P-labeled (specific activity of 25 and 4.5 X 10⁸ dpm/μg DNA, respectively). The positions of molecular size markers are indicated on the left, and the linear form of pFN1, pFN2 and pFN3 are indicated on the right.

Figure 2A shows physical characteristics of plasmid pFN1 based on DNA sequence analysis. Open reading frames (ORFs), the putative origin of replication (*ori*), and selected restriction endonuclease sites are indicated.

Figure 2B shows structural elements of putative origin of replication found upstream of ORF5, the *repA* homologue. The putative origin contains an A-T rich region (cross hatched bar), six perfect 22 bp direct repeats (▲) termed iterons, and several putative DnaA-binding sites (■).

Figure 3 shows another partial restriction maps of plasmids pFN1 and pFN2. Restriction endonuclease sites of selected enzymes are indicated.

Figure 4 shows Southern blot analysis of pFN1 and pFN2 DNA with the pFN1 *repA* and *rlx* gene probes. The left panel shows agarose gel demonstrating restriction endonuclease digestion fragments of pFN1 and pFN2 DNA. The middle panel shows Southern blot of pFN1 and pFN2 DNA with pFN1 *repA* gene probe. The right panel shows Southern blot of pFN1 and pFN2 DNA with pFN1 *rlx* gene probe. For all panels: lane 1, *EcoRI* digested pFN1; lane 2, *StyI* digested pFN1; lane 3, *EcoRI* digested pFN2; lane 4, *HincII* digested pFN2; lane 5, *HincII/HpaII* digested pFN2; lane 6, *AvrII* digested pFN2.

Figure 5 shows localization of *repA* and *rlx* homology on pFN2. Restriction maps of pFN1 and pFN2 indicate the positions of selected restriction

endonuclease sites, the repA and rlx genes of pFN1, and the repA and rlx homologous regions of pFN2.

Figure 6 shows that plasmid DNA from *F. nucleatum* ATCC 10953 transformants consists of the shuttle plasmid, pHS17, and the native plasmid, pFN3.

5 Plasmid preparations from *E. coli* (pHS17), *F. nucleatum* ATCC 10953 transformant strain KH21 (pHS17 and pFN3) and *F. nucleatum* ATCC 10953 (pFN3) were analyzed. The preparations were not digested (lanes 1), or predigested with *EcoRV* (lanes 2) or *EcoRI* (lanes 3), separated on 0.8% agarose gels, stained with ethidium bromide and visualized under UV illumination. The open circular (OC), linear (L) and covalently
10 closed circular (CC) forms of pHS17 and pFN3 are indicated on the left and right, respectively.

DEFINITIONS

As used herein, the following terms have the meanings ascribed to them
15 unless specified otherwise.

The terms "isolated," "purified," or "biologically pure" refer to material that is substantially or essentially free from components which normally accompany it as found in its native state. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high
20 performance liquid chromatography. A protein that is the predominant species present in a preparation is substantially purified. For example, an isolated repA nucleic acid is separated from open reading frames that flank the repA gene and encode proteins other than a RepA protein. The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic
25 acid or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

"Nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or
30 linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs).

Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. The term nucleic acid is used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an analog or mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. Polypeptides can be modified, *e.g.*, by the addition of carbohydrate residues to form glycoproteins. The terms "polypeptide," "peptide" and "protein" include glycoproteins, as well as non-glycoproteins.

The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline, carboxyglutamate, and O-phosphoserine. Amino acid analogs refer to compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, a carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group. Exemplary amino acid analogs include, *e.g.*, homoserine, norleucine, methionine sulfoxide, and methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refer to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that function in a manner similar to a naturally occurring amino acid.

Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes (A, T, G, C, U, *etc.*).

"Conservatively modified variations" of a particular polynucleotide sequence refers to those polynucleotides that encode identical or essentially identical amino acid sequences, or where the polynucleotide does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic

code, a large number of functionally identical nucleic acids encode any given polypeptide. For instance, the codons CGU, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at every position where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent substitutions" or "silent variations," which are one species of "conservatively modified variations." Every polynucleotide sequence described herein which encodes a polypeptide also describes every possible silent variation, except where otherwise noted. Thus, silent substitutions are an implied feature of every nucleic acid sequence which encodes an amino acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule by standard techniques. In some embodiments, the nucleotide sequences that encode the enzymes are preferably optimized for expression in a particular host cell (*e.g.*, yeast, mammalian, plant, fungal, and the like) used to produce the enzymes.

Similarly, "conservative amino acid substitutions," in one or a few amino acids in an amino acid sequence are substituted with different amino acids with highly similar properties are also readily identified as being highly similar to a particular amino acid sequence, or to a particular nucleic acid sequence which encodes an amino acid. Such conservatively substituted variations of any particular sequence are a feature of the present invention. Individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 1%) in an encoded sequence are "conservatively modified variations" where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. For example, the following groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Glycine (G);
- 2) Serine (S), Threonine (T);
- 3) Aspartic acid (D), Glutamic acid (E);
- 4) Asparagine (N), Glutamine (Q);
- 5) Cysteine (C), Methionine (M);
- 6) Arginine (R), Lysine (K), Histidine (H);
- 7) Isoleucine (I), Leucine (L), Valine (V); and
- 8) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

(see, e.g., Creighton, *Proteins*, W.H. Freeman and Company (1984)).

A "label" is a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include ^{32}P , fluorescent dyes, electron-dense reagents, enzymes (e.g., as commonly used in an ELISA), biotin, digoxigenin, or haptens and proteins for which antisera or monoclonal antibodies are available (e.g., the polypeptide of SEQ ID NO:1 can be made detectable, e.g., by incorporating a radiolabel into the peptide, and used to detect antibodies specifically reactive with the peptide).

As used herein a "nucleic acid probe or oligonucleotide" is defined as a nucleic acid capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. As used herein, a probe may include natural (*i.e.*, A, G, C, or T) or modified bases (7-deazaguanosine, inosine, *etc.*). In addition, the bases in a probe may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization. Thus, for example, probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages. It will be understood by one of skill in the art that probes may bind target sequences lacking complete complementarity with the probe sequence depending upon the stringency of the hybridization conditions. The probes are preferably directly labeled as with isotopes, chromophores, lumiphores, chromogens, or indirectly labeled such as with biotin to which a streptavidin complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of the selected sequence or subsequence.

A "labeled nucleic acid probe or oligonucleotide" is one that is bound, either covalently, through a linker or a chemical bond, or noncovalently, through ionic, van der Waals, electrostatic, or hydrogen bonds to a label such that the presence of the probe may be detected by detecting the presence of the label bound to the probe.

The term "recombinant" when used with reference, e.g., to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

A "vector" refers to a carrier DNA molecule into which a nucleic acid sequence can be inserted for introduction into a new host cell where it will be replicated, and in some cases expressed. Vectors can be derived from plasmids, bacteriophages, plant, animal viruses, *etc.*

5 An "expression vector" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell. Typically, the expression vector includes a nucleic acid to be transcribed operably linked to a promoter.

10 The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (*i.e.*, 70% identity, optionally 75%, 80%, 85%, 90%, 95%, 97%, 98%, or 99%) identity over a specified region), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one
15 of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be "substantially identical." This definition also refers to the complement of a test sequence. Optionally, the identity exists over a region that is at least about 50 amino acids or nucleotides in length, preferably over a region that is at least about 75 amino acids or nucleotides in length, and most preferably
20 over a region that is at least about 100 amino acids or nucleotides in length. In most preferred embodiments, the sequences are substantially identical over the entire length of, *e.g.*, the coding region.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison
25 algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program
30 parameters.

A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of

contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (*see, e.g., Current Protocols in Molecular Biology* (Ausubel *et al.*, eds. 1995 supplement)).

Preferred examples of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.*, *Nuc. Acids Res.* 25:3389-3402 (1997) and Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990), respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix

(see, Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

A further indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below. Yet another indication that two nucleic acid sequences are substantially identical is that the same primers can be used to amplify the sequence.

The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10 °C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The

T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m , 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt

5 concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60°C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific
10 hybridization, a positive signal is at least two times background, preferably 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5x SSC, and 1% SDS, incubating at 42°C, or 5x SSC, 1% SDS, incubating at 65°C, with a wash in 0.2x SSC, and 0.1% SDS at 65°C.

Nucleic acids that do not hybridize to each other under stringent conditions
15 are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions. Exemplary "moderately stringent hybridization conditions" include a hybridization in a buffer of
20 40% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 1X SSC at 45°C. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

"Antibody" refers to a polypeptide substantially encoded by an
25 immunoglobulin gene or immunoglobulin genes, or fragments thereof, which specifically binds and recognizes an epitope (*e.g.*, an antigen). The recognized immunoglobulin genes include the kappa and lambda light chain constant region genes, the alpha, gamma, delta, epsilon and mu heavy chain constant region genes, and the myriad immunoglobulin variable region genes. Antibodies exist, *e.g.*, as intact immunoglobulins or as a number
30 of well characterized fragments produced by digestion with various peptidases. This includes, *e.g.*, Fab' and F(ab)'₂ fragments. The term "antibody," as used herein, also includes antibody fragments either produced by the modification of whole antibodies or those synthesized *de novo* using recombinant DNA methodologies. It also includes

polyclonal antibodies, monoclonal antibodies, chimeric antibodies, humanized antibodies, or single chain antibodies. "Fc" portion of an antibody refers to that portion of an immunoglobulin heavy chain that comprises one or more heavy chain constant region domains, CH₁, CH₂ and CH₃, but does not include the heavy chain variable region.

5 An "anti-RepA antibody" is an antibody or antibody fragment that specifically binds a polypeptide encoded by a repA gene, cDNA, or a subsequence thereof.

 The term "immunoassay" is an assay that uses an antibody to specifically bind an antigen. The immunoassay is characterized by the use of specific binding
10 properties of a particular antibody to isolate, target, and/or quantify the antigen.

 The phrase "specifically (or selectively) binds" to an antibody or "specifically (or selectively) immunoreactive with," when referring to a protein or peptide, refers to a binding reaction that is determinative of the presence of the protein in a heterogeneous population of proteins and other biologics. Thus, under designated
15 immunoassay conditions, the specified antibodies bind to a particular protein at least two times the background and do not substantially bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, polyclonal antibodies raised to a RepA protein with the amino acid sequence of
20 a region encoded in SEQ ID NO:1 can be selected to obtain only those polyclonal antibodies that are specifically immunoreactive with the region of the RepA protein and not with other proteins, except for, e.g., homologs or variants of the RepA protein. This selection may be achieved by subtracting out antibodies that cross react with molecules such as other RepA proteins. A variety of immunoassay formats may be used to select
25 antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select antibodies specifically immunoreactive with a protein (see, e.g., Harlow & Lane, *Antibodies, A Laboratory Manual* (1988), for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity). Typically a specific or selective reaction will be
30 at least twice background signal or noise and more typically more than 10 to 100 times background.

 The phrase "selectively associates with" refers to the ability of a nucleic acid to "selectively hybridize" with another as defined above, or the ability of an antibody to "selectively (or specifically) bind to a protein, as defined above.

The term "host cell" is meant that a cell that contains a plasmid or a vector and supports the replication or expression of the plasmid or the vector. Host cells include, but are not limited to, *E. coli*, *F. nucleatum*, *Leptotrichia*, *Streptococcus*, *Staphylococcus*, *Clostridium*, etc.

5

THE DETAILED DESCRIPTION OF THE INVENTION

The invention provides novel plasmid origin of replication sequences and replication genes and proteins (hereinafter referred to as "repA" for nucleic acids and "RepA" for polypeptides) that are functional in *Fusobacterium* (e.g., *F. nucleatum*). Also provided by the invention are plasmids and vectors that can replicate in *Fusobacterium* and related species. In some embodiment, the plasmids and vectors of the invention comprise additional origin of replication sequences so that they can replicate in other microorganisms, such as *E. coli*. Further, the invention provides host cells and methods of transforming the host cells using the plasmids and vectors of the present invention.

10

15

The above-described embodiments of the invention are useful for several purposes. The origin of replication sequences and repA nucleic acids can be used to construct cloning and expression plasmids and vectors that are functional in *Fusobacterium* as well as in other microorganisms, such as *E. coli*. The plasmids and vectors of the invention can be used, e.g., for cloning of *F. nucleatum* nucleic acids as well as expression of these genes in the native host background. The plasmids and vectors can also be used for cloning and/or expression of foreign genes. Moreover, the plasmids and vectors of the invention can be used in the development of vaccine delivery systems. For example, foreign genes that encode, e.g., an antigenic determinant of a pathogenic organism can be introduced into the plasmids or vectors of the invention. *F. nucleatum* transformed with these plasmids or vectors can be introduced into the oral cavity as a vaccine delivery system to stimulate mucosal immunity against these pathogens.

20

25

I. Isolation of Plasmid Origin of Replication Sequences and repA Nucleic Acids

30

A. Origin of Replication Sequences for *Fusobacterium* Plasmids

In one aspect, the invention provides origin of replication sequences of a plasmid functional in *Fusobacterium*. An origin of replication in a plasmid provides a region at which specific proteins bind to the open DNA complex, thereby initiating the

plasmid DNA replication process. The origin of replication sequences of the invention are useful for several purposes. For example, the origin of replication sequences of the invention are useful as probes to identify other plasmid origin of replication sequences. In another example, the origin of replication sequences can be used to construct a plasmid or a vector that can replicate in *Fusobacterium* (e.g., *F. nucleatum*). The origin of replication sequences can also be used to construct a shuttle vector that can replicate in *Fusobacterium* as well as in other microorganisms, such as *E. coli*.

An example of an origin of replication sequences of the invention is isolated from *Fusobacterium* species, such as *F. nucleatum*. A presently preferred origin of replication sequence comprises a nucleic acid as shown in SEQ ID NO:4 and is derived from plasmid pFN1 of *F. nucleatum* strain 12230 (a gift from S. Finegold, West Los Angeles, VA). As shown in the Sequence Listing, SEQ ID NO:4 contains six perfect 22 base pair direct repeats ("iterons"). The 22 base pair iteron is given SEQ ID NO:3.

The origin of replication sequences of the present invention generally include those that comprise at least two copies of the iteron, wherein the iteron has the nucleic acid sequence of SEQ ID NO:3. Preferably, the origin of replication sequences of the invention comprise between two to six copies of the iteron having the nucleic acid sequence of SEQ ID NO:3. More preferably, the origin of replication sequences of the invention comprise a nucleic acid sequence of SEQ ID NO:4 (i.e., six copies of the iteron having SEQ ID NO:3, which six copies span the nucleotide positions 4169 to 4300 of plasmid pFN1). In some embodiments, the origin of replication sequences can comprise imperfect repeats of the iteron having SEQ ID NO:3. For example, the origin of replication sequences can comprise nucleic acids as shown in SEQ ID NOs: 11, 12, 13, or any combinations thereof. Moreover, the origin of replication can further comprise other sequences, such as DnaA binding site sequences and A-T rich regions as shown in Figure 3B. For instance, in some embodiments the origin of replication sequences can comprise the nucleotide position 3936 to 4481 of plasmid pFN1. This fragment of plasmid pFN1 comprises six copies of the iterons as well as the DnaA binding site sequences and A-T rich regions.

To identify origin of replication sequences of the invention, one can use visual inspection or can use a suitable algorithm as described herein. Alternatively, one can identify origin of replication sequences of the invention by hybridizing, under stringent conditions, the candidate nucleic acids to the origin of replication sequences (e.g., a sequence comprising SEQ ID NO:4).

B. repA Nucleic Acids for *Fusobacterium* Plasmids

In another aspect, the invention provides repA nucleic acids encoding RepA polypeptides that can bind to origin of replication sequences in a plasmid of *Fusobacterium*. Not wishing to be bound by a theory, the RepA polypeptides encoded by the repA nucleic acids bind to the iteron sequences in the origin of replication of a plasmid. The binding results in structural changes, including melting of the adjacent A-T rich region, to form an open complex. The RepA polypeptides, possibly in conjunction with the host DnaA protein, then guide host replication proteins into the open complex, and thereby initiating the plasmid DNA replication process.

The repA nucleic acids of the invention are useful for several purposes. For example, repA nucleic acids can be used to recombinantly express RepA polypeptides. These RepA polypeptides can then be used as immunogens to produce anti-RepA polypeptide antibodies. The repA nucleic acids can also be used as probes to identify other repA nucleic acids. Moreover, the repA nucleic acids of the invention can be used to construct a plasmid or a vector functional in *Fusobacterium*.

An example of a repA nucleic acid of the invention is produced by *Fusobacterium* species, such as *F. nucleatum*. A presently preferred repA nucleic acid is that of plasmid pFN1, which has a nucleic acid sequence as shown in SEQ ID NO:2 and encodes a polypeptide of 407 amino acids as shown in SEQ ID NO:1. Plasmid pFN1 can be isolated from *F. nucleatum* strain 12230 (a gift from S. Finegold, West Los Angeles, VA). The repA nucleic acid is located at nucleotide position 4508 to 5731 of plasmid pFN1, and contains putative promoters at -35 and -10 nucleotide positions.

The repA nucleic acids of the invention generally include those that encode a RepA polypeptide having an amino acid sequence that is at least about 80% identical to an amino acid sequence as set forth in SEQ ID NO:1 over a region of at least about 50 amino acids in length. Preferably, the RepA polypeptides encoded by the nucleic acids of the invention are at least about 85% identical to an amino acid sequence of SEQ ID NO:1, more preferably are at least about 90% identical to an amino acid sequence of SEQ ID NO:1, still more preferably are at least about 95% identical to an amino acid sequence of SEQ ID NO:1, and most preferably are at least about 96%, 97%, 98%, or 99% identical to an amino acid sequence of SEQ ID NO:1, over a region of at least 50 amino acids in length. In preferred embodiments, the region of percent identity extends over a region longer than 50 amino acids, preferably over a region of at least

about 75 amino acids, more preferably over a region of at least about 100 amino acids, and most preferably over the full length of the RepA polypeptide. In a preferred embodiment, the repA nucleic acids of the invention encode a polypeptide having an amino acid sequence as shown in SEQ ID NO:1.

Moreover, the repA nucleic acids of the invention are typically at least about 80% identical to a nucleic acid sequence of SEQ ID NO:2 over a region of at least about 50 nucleotides in length. Preferably, the repA nucleic acids of the invention are at least about 85% identical to a nucleic acid of SEQ ID NO:2, more preferably are at least about 90% identical to a nucleic acid of SEQ ID NO:2, and most preferably are at least about 95% identical to a nucleic acid of SEQ ID NO:2, over a region of at least about 50 nucleotides in length. In preferred embodiments, the region of percent identity extends over a region longer than 50 nucleotides, preferably over a region of at least about 75 nucleotides, more preferably over a region of at least about 100 nucleotides, and most preferably over the full length of the RepA encoding region.

To identify repA nucleic acids of the invention, one can use visual inspection or can use a suitable alignment algorithm as described herein, such as the BLASTN Version 2.0 algorithm. An alternative method by which one can identify a repA nucleic acid of the invention is by hybridizing, under stringent conditions, the candidate repA nucleic acids to the repA nucleic acids described herein (*e.g.*, a sequence comprising a sequence of SEQ ID NO:2).

C. Methods for Isolation of Nucleic Acids

The origin of replication sequences or repA nucleic acids of the invention can be obtained using methods that are known to those of skill in the art. Suitable nucleic acids (*e.g.*, cDNA, plasmid, or subsequences (probes)) can be cloned, or amplified by *in vitro* methods such as the polymerase chain reaction (PCR), the ligase chain reaction (LCR), the transcription-based amplification system (TAS), the self-sustained sequence replication system (SSR). A wide variety of cloning and *in vitro* amplification methodologies are well-known to persons of skill. Examples of these techniques and instructions sufficient to direct persons of skill through many cloning exercises are found in Berger & Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology* 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual* (2nd ed.) Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, NY (1989); *Current Protocols in Molecular Biology*, Ausubel *et al.*, eds.,

Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1994 Supplement); Cashion *et al.*, U.S. Patent No. 5,017,478; and Carr, European Patent No. 0,246,864. Examples of techniques sufficient to direct persons of skill through *in vitro* amplification methods are found in Berger, Sambrook, and Ausubel, as well as Mullis *et al.*, (1987) U.S. Patent No. 4,683,202; *PCR Protocols A Guide to Methods and Applications* (Innis *et al.*, eds) Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim & Levinson *C&EN* 36-47 (October 1, 1990); *The Journal Of NIH Research* 3: 81-94 (1991); Kwoh *et al. Proc. Natl. Acad. Sci. USA* 86: 1173 (1989); Guatelli *et al.*, *Proc. Natl. Acad. Sci. USA* 87 (1990), 1874; Lomell *et al.*, *J. Clin. Chem.*, 35: 1826 (1989); Landegren *et al.*, *Science* 241: 1077-1080 (1988); Van Brunt, *Biotechnology* 8: 291-294 (1990); Wu & Wallace, *Gene* 4: 560 (1989); and Barringer *et al.*, *Gene* 89: 117 (1990). Improved methods of cloning *in vitro* amplified nucleic acids are described in Wallace *et al.*, U.S. Patent No. 5,426,039.

Origin of replication sequences or repA nucleic acids of the invention, or subsequences thereof can be obtained using any suitable method as described above, including, for example, cloning and restriction of appropriate sequences. In cloning methods, a known nucleotide sequence of an origin of replication, such as those described herein, can be used to provide probes that specifically hybridize to other nucleic acids that encode an origin of replication in a plasmid DNA samples. Similarly, a known nucleotide sequence of a repA gene, such as those described herein, can be used to provide probes that specifically hybridize to a gene that encodes a RepA polypeptide in a plasmid DNA sample, or to a mRNA in a total RNA sample (*e.g.*, in a Southern or Northern blot). Preferably, the samples are obtained from prokaryotic organisms, such as *Fusobacterium* species. Examples of *Fusobacterium* species of particular interest include *F. nucleatum*, *F. necrophorum*, *F. varium*, *F. periodonticum*, *etc.*

Once the target nucleic acid is identified, it can be isolated according to standard methods known to those of skill in the art (*see, e.g.*, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual, 2nd Ed., Vols. 1-3*, Cold Spring Harbor Laboratory (1989); Berger and Kimmel, *Methods in Enzymology, Vol. 152: Guide to Molecular Cloning Techniques*, San Diego: Academic Press, Inc. (1987); or Ausubel *et al.*, *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York (1987).

The origin of replication sequences, repA nucleic acids, or subsequences thereof can also be cloned using DNA amplification methods such as polymerase chain

reaction (PCR). For example, the nucleic acid sequence or subsequence of an origin of replication or a repA gene is PCR amplified, preferably using a sense primer containing one restriction site (*e.g.*, *Xba*I) and an antisense primer containing another restriction site (*e.g.*, *Hind*III). This will produce an origin of replication nucleic acid, a repA nucleic acid, or a subsequence thereof, having terminal restriction sites. This nucleic acid can then be ligated into a vector containing a nucleic acid encoding the second molecule and having the appropriate corresponding restriction sites. Suitable PCR primers can be determined by one of skill in the art using the sequence information provided herein. Examples of suitable primers for amplification of the origin of replication sequences or repA nucleic acids are provided in Example 4. Appropriate restriction sites can also be added to the origin of replication sequences or repA nucleic acids, by site-directed mutagenesis. The plasmid containing the origin of replication sequences, repA nucleic acids, and/or subsequences thereof is cleaved with the appropriate restriction endonuclease and then ligated into an appropriate vector for amplification and/or expression according to standard methods.

Since repA nucleic acids encode detectable polypeptides, these nucleic acids can also be cloned by detecting their expressed RepA polypeptides using assays based on, *e.g.*, immunological properties. For example, one can identify a cloned repA nucleic acid by screening expression libraries using antibodies as probes. Such polyclonal or monoclonal antibodies can be raised using a sequence or a subsequence of, *e.g.*, SEQ ID NO:1.

As an alternative to cloning the origin of replication sequences or repA nucleic acids, a suitable nucleic acid can be chemically synthesized from known sequences of origin of replication or repA of the invention (*e.g.*, SEQ ID NO:4 or SEQ ID NO:2, respectively). Direct chemical synthesis methods include, for example, the phosphotriester method of Narang *et al.*, *Meth. Enzymol.* 68: 90-99 (1979); the phosphodiester method of Brown *et al.*, *Meth. Enzymol.* 68: 109-151 (1979); the diethylphosphoramidite method of Beaucage *et al.*, *Tetra. Lett.*, 22: 1859-1862 (1981); and the solid support method of U.S. Patent No. 4,458,066. Chemical synthesis produces a single stranded oligonucleotide. This can be converted into double stranded DNA by hybridization with a complementary sequence, or by polymerization with a DNA polymerase using the single strand as a template. One of skill would recognize that while chemical synthesis of DNA is often limited to sequences of about 100 bases, longer sequences may be obtained by the ligation of shorter sequences. Alternatively,

subsequences may be cloned and the appropriate subsequences cleaved using appropriate restriction enzymes. The fragments can then be ligated to produce the desired DNA sequence.

In some embodiments, it may be desirable to modify the origin of replication sequences or repA nucleic acids. One of skill will recognize many ways of generating alterations in a given nucleic acid construct. Such well-known methods include site-directed mutagenesis, PCR amplification using degenerate oligonucleotides, exposure of cells containing the nucleic acid to mutagenic agents or radiation, chemical synthesis of a desired oligonucleotide (e.g., in conjunction with ligation and/or cloning to generate large nucleic acids) and other well-known techniques. See, e.g., Gilman & Smith, *Gene* 8:81-97 (1979); Roberts *et al.*, *Nature* 328: 731-734 (1987).

II. Isolation of RepA Polypeptides

In yet another aspect, the invention provides RepA polypeptides that bind to an origin of replication sequence of a plasmid in *Fusobacterium*. An example of a RepA polypeptide of the invention is produced by *Fusobacterium* species, such as *F. nucleatum*. A presently preferred RepA polypeptide has an amino acid sequence as shown in SEQ ID NO:1 (consisting of 407 amino acids) and has an approximate molecular weight of 44.8 kDa. The RepA polypeptides of the invention can be purified from natural sources, e.g., *F. nucleatum* strain 12230 (a gift from S. Finegold, West Los Angeles), or can be recombinantly made. Methods for recombinant production or purification of RepA polypeptides are well known in the art. Substantially pure compositions of at least about 90 to 95% homogeneity are preferred for some applications, and 98 to 99% or more homogeneity are most preferred. Once purified, partially or to homogeneity as desired, the RepA polypeptides may then be used, e.g., as immunogens for antibody production.

The RepA polypeptides of the invention generally include an amino acid sequence that is at least about 80% identical to an amino acid sequence as set forth in SEQ ID NO:1 over a region at least about 50 amino acids in length. Preferably, the RepA polypeptides of the invention are at least about 85% identical to the amino acid sequence of SEQ ID NO:1, more preferably are at least about 90% identical to the amino acid sequence of SEQ ID NO:1, still more preferably are at least about 95% identical to the amino acid of SEQ ID NO:1, and most preferably are at least about 96%, 97%, 98%, 99% identical to the amino acid of SEQ ID NO:1, over a region of 50 amino acids in length.

In presently preferred embodiments, the region of percent identity extends over a region of at least about 50 amino acids, preferably over a region of at least about 75 amino acids, more preferably over a region of at least about 100 amino acids, and most preferably over the full length of the RepA polypeptide.

5 To identify RepA polypeptides of the invention, one can use visual inspection or can use a suitable alignment algorithm as described above, such as the BLASTP Version 2.0 algorithm. Alternatively, the RepA polypeptides of the invention can also be identified by immunoreactivity. For example, one can produce RepA antibodies against an antigenic determinant of the RepA polypeptide having SEQ ID
10 NO:1 and determine whether the antibodies are specifically immunoreactive with a RepA polypeptide of interest.

002221 3324260
15 The RepA polypeptides of the invention can be isolated from natural sources (e.g., from *F. nucleatum*) or synthesized using the recombinant nucleic acid or chemical synthesis methodologies. These methodologies are well known in the art. If recombinant expression is desired, the repA nucleic acid sequences of the invention described above can be operably linked to appropriate control sequences for expression in a host cell (e.g., bacterial, yeast, plant, fungi, or mammalian cells). For example, if *E. coli* or *F. nucleatum* is used as a host cell, the repA nucleic acid sequences are operably
20 linked to a promoter, a ribosome binding site and preferably, a transcription termination signal. The repA nucleic acid sequence operably linked to control sequences is introduced into an appropriate plasmid or a vector for expression of RepA polypeptides in a host cell.

25 Once expressed, the naturally occurring or recombinant RepA polypeptides can be purified according to standard procedures of the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis and the like (see, e.g., Scopes, *Polypeptide Purification* (1982); Deutscher, *Methods in Enzymology Vol. 182: Guide to Polypeptide Purification* (1990)).

30 If chemical synthesis is desired, the RepA polypeptides can be synthetically prepared via a wide variety of well-known techniques. Polypeptides of relatively short size are typically synthesized in solution or on a solid support in accordance with conventional techniques (see, e.g., Merrifield, *Am. Chem. Soc.* 85:2149-2154 (1963)). Various automatic synthesizers and sequencers are commercially available and can be used in accordance with known protocols (see, e.g., Stewart & Young, *Solid Phase Peptide Synthesis* (2nd ed. 1984)). Solid phase synthesis in which

the C-terminal amino acid of the sequence is attached to an insoluble support followed by sequential addition of the remaining amino acids in the sequence is the preferred method for the chemical synthesis of the polypeptides of this invention. Techniques for solid phase synthesis are described by Barany & Merrifield, *Solid-Phase Peptide Synthesis*; pp. 3-284 in *The Peptides: Analysis, Synthesis, Biology. Vol. 2: Special Methods in Peptide Synthesis, Part A.*; Merrifield *et al.*, *J. Am. Chem. Soc.* 85:2149-2156 (1963); and Stewart *et al.*, *Solid Phase Peptide Synthesis* (2nd ed. 1984).

After chemical synthesis, biological expression or purification, the RepA polypeptide may possess a conformation substantially different than the native conformations of the constituent polypeptides. In this case, it is helpful to denature and reduce the RepA polypeptide and then to cause the polypeptide to re-fold into the preferred conformation. Methods of reducing and denaturing polypeptides and inducing re-folding are well known to those of skill in the art (*see*, Debinski *et al.*, *J. Biol. Chem.* 268:14065-14070 (1993); Kreitman & Pastan, *Bioconjug. Chem.* 4:581-585 (1993); and Buchner *et al.*, *Anal. Biochem.* 205:263-270 (1992)). Debinski *et al.*, for example, describe the denaturation and reduction of inclusion body polypeptides in guanidine-DTE. The polypeptide is then refolded in a redox buffer containing oxidized glutathione and L-arginine.

One of skill will recognize that modifications can be made to the RepA polypeptides without diminishing their biological activity. Some modifications may be made to facilitate the cloning, expression, or incorporation of the targeting molecule into a fusion polypeptide. Such modifications are well known to those of skill in the art and include, for example, a methionine added at the amino terminus to provide an initiation site, or additional amino acids placed on either terminus to create conveniently located restriction sites or termination codons or purification sequences.

III. Immunological Detection of RepA Polypeptides

In addition to the detection of repA genes using the nucleic acid hybridization technology, one can also use immunoassays to detect RepA polypeptides. Immunoassays can be used to qualitatively or quantitatively analyze RepA polypeptides. A general overview of the applicable technology can be found in Harlow & Lane, *Antibodies: A Laboratory Manual* (1988). *See, also*, U.S. Patent Nos. 4,366,241; 4,376,110; 4,517,288; and 4,837,168. Useful assays include, for example, an enzyme immune assay (EIA) such as enzyme-linked immunosorbent assay (ELISA), a

radioimmune assay (RIA), a Western blot assay, or a slot blot assay. For a review of the general immunoassays, *see also, Methods in Cell Biology: Antibodies in Cell Biology*, volume 37 (Asai, ed. 1993); *Basic and Clinical Immunology* (Stites & Terr, eds., 7th ed. 1991).

5 Antibodies that specifically bind to a RepA polypeptide can be prepared using any suitable methods known in the art. *See, e.g., Coligan, Current Protocols in Immunology* (1991); Harlow & Lane, *supra*; Goding, *Monoclonal Antibodies: Principles and Practice* (2d ed. 1986); and Kohler & Milstein, *Nature* 256:495-497 (1975). Such techniques include antibody preparation by selection of antibodies from libraries of
10 recombinant antibodies in phage or similar vectors, as well as preparation of polyclonal and monoclonal antibodies by immunizing rabbits or mice (*see, e.g., Huse et al., Science* 246:1275-1281 (1989); Ward *et al., Nature* 341:544-546 (1989)). Specific polyclonal antisera and monoclonal antibodies will usually bind with a K_d of at least about 0.1 mM, more usually at least about 1 μ M, preferably at least about 0.1 μ M or better, and most
15 preferably, 0.01 μ M or better.

 After the antibody is provided, a sample comprising the RepA polypeptides can be contacted with the antibody. Optionally, the antibody can be fixed to a solid support to facilitate washing and subsequent isolation of the complex, prior to contacting the antibody with a sample. Examples of solid supports include glass or
20 plastic in the form of, *e.g.,* a microtiter plate, a stick, a bead, or a microbead. After contacting the sample with the antibody, the mixture is incubated for 10 seconds to 12 hours, preferably from about 30 seconds to about 30 minutes.

 After incubation, the mixture is washed and the presence or amount of antibody-RepA polypeptide complex formed is determined. This can be accomplished by
25 incubating the washed mixture with a detection reagent (*e.g.,* a second, labeled antibody). This detection reagent may be a monoclonal or polyclonal antibody and is labeled with a detectable label. Exemplary detectable labels include magnetic beads (*e.g.,* DYNABEADSTM), fluorescent dyes, radiolabels, enzymes (*e.g.,* horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels
30 such as colloidal gold or colored glass or plastic beads. Alternatively, the presence or amount of RepA polypeptides in the sample can be determined using an indirect assay, wherein, for example, a second, labeled antibody is used to detect bound antibody, and/or in a competition or inhibition assay wherein, for example, a monoclonal antibody which

binds to a distinct epitope of the marker are incubated simultaneously with the mixture. These techniques are well-known in the art and are within the skill of those in the art.

To determine if a protein specifically binds to the polyclonal antibodies generated to RepA polypeptides or fragments thereof, immunoassays in the competitive binding format can be used. For example, a protein having an amino acid sequence of SEQ ID NO:1 or a fragment thereof can be immobilized to a solid support. Other proteins (candidate RepA polypeptides) are added to the assay so as to compete for binding of antisera to the immobilized antigen. The ability of the added proteins to compete for binding of the antisera to the immobilized protein is compared to the ability of the RepA polypeptide having SEQ ID NO:1 or a fragment thereof compete with itself. The percent crossreactivity of the above proteins is calculated, using standard calculations. Those antisera with less than 10% crossreactivity with each of the added proteins listed above are selected and pooled. The cross-reacting antibodies are optionally removed from the pooled antisera by immunoabsorption with the added proteins, e.g., distantly related homologs.

The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay to compare a second protein, thought to be a homolog of RepA polypeptide, to the immunogen protein. In order to make this comparison, the two proteins are each assayed at a wide range of concentrations and the amount of each protein required to inhibit 50% of the binding of the antisera to the immobilized protein is determined. If the amount of the second protein required to inhibit 50% of binding is less than 10 times the amount of the protein having an amino acid sequence of SEQ ID NO:1 or a fragment thereof that is required to inhibit 50% of binding, then the second protein is said to specifically bind to the polyclonal antibodies generated to the RepA polypeptide of the invention.

IV. Isolation of *Fusobacterium* Plasmids

In yet another aspect, the invention provides plasmids that can be stably maintained in *Fusobacterium*. The plasmids of the invention include, for example, those that are isolated from the natural source (e.g., *F. nucleatum*), modified plasmids (e.g., have substitution, deletion or addition of sequences) or recombinant plasmids. These plasmids are useful for several purposes. For example, they can be used as a cloning or expression system for *Fusobacterium*. The plasmids of the invention can also be used to

construct a shuttle vector that can replicate in *Fusobacterium*, as well as in another microorganism, such as *E. coli*.

The plasmids of the invention can be purified from natural sources, *e.g.*, *Fusobacterium* species described herein. In preferred embodiments, plasmids are obtained from *F. nucleatum*. Methods by which the plasmids can be isolated include standard plasmid DNA miniprep methods including, *e.g.*, the alkaline lysis prep, the boiling methods and a lithium-based procedure (*see, e.g.*, Ausubel *et al.*, *supra.*). Alternatively, the plasmids of the invention can be recombinantly produced. For example, plasmid pFN1 or its derivative can be produced recombinantly using the sequence information disclosed herein. A number of cloning and *in vitro* amplification methodologies described above can also be used to produce plasmids of the invention.

Embodiments of the present plasmids can be isolated from *Fusobacterium* species, such as *F. nucleatum*. These plasmids include, *e.g.*, pFN1, pFN2 and pFN3. Plasmid pFN1 can be isolated from, *e.g.*, *F. nucleatum* strain 12230 (a gift from S. Finegold, West Los Angeles, VA). Plasmid pFN1 is characterized by a length of about 5.9 kb (*see*, GenBank Accession No. AF159249) and has partial restriction maps as shown in Figures 1A, 2, 3 and 5. For the partial restriction map shown in Figure 1A, the number of cleavage sites for each restriction enzyme and the length of fragments generated by the restriction enzymes for pFN1 are as follows:

TABLE 1

Restriction Enzymes	Number of Cleavage Sites	DNA Fragments (kb)
<i>HincII</i>	3	1.9, 2.9, 1.1
<i>Fnu4HI</i>	2	2.3, 3.6
<i>HindIII</i>	3	1.5, 0.1, 4.3
<i>AvrII</i>	1	5.9
<i>EcoRI</i>	1	5.9

Plasmid pFN2 can be isolated from, *e.g.*, *F. nucleatum* strain 10113 (a gift from S. Finegold, West Los Angeles, VA). Plasmid pFN2 is characterized by a length of about 7.2 kb and has partial restriction maps as shown in Figure 1A, 3 and 5. For the partial restriction map shown in Figure 1A, the number of cleavage sites for each restriction enzyme and the length of fragments generated by the restriction enzymes for pFN2 are as follows:

TABLE 2

Restriction Enzymes	Number of Cleavages	DNA Fragments (kb)
<i>SpeI</i>	1	7.2
<i>NdeI</i>	1	7.2
<i>HincII</i>	3	1.0, 2.8, 3.4
<i>Fnu4HI</i>	2	3.8, 3.4
<i>HindIII</i>	2	5.4, 1.8
<i>AvrII</i>	2	5.5, 1.7
<i>EcoRI</i>	1	7.2
<i>BspHI</i>	1	7.2

Plasmid pFN3 can be isolated from, *e.g.*, *F. nucleatum* strain ATCC Accession No. 10953. Plasmid pFN3 is characterized by a length of about 11.1 kb and has a partial restriction map as shown in Figure 1A. The nucleotide sequence of plasmid pFN3 has been partially determined and is shown as SEQ ID NO: 14. The number of cleavage sites for each restriction enzyme and the length of fragments generated by the restriction enzymes for pFN3 are as follows:

TABLE 3

Restriction Enzyme	Number of Cleavage Sites	DNA Fragments (kb)
<i>HindIII</i>	2	3.4, 7.7
<i>EcoRV</i>	1	11.1
<i>AflIII</i>	2	2.1, 9.0

Each of these plasmids comprises an origin of replication and a *repA* nucleic acid which allows them to replicate in *Fusobacterium*. For example, plasmid pFN1 comprises an origin of replication and *repA* nucleic acid sequences within the 2.36 kb restriction segment between restriction sites *AvrII* and *ScaI* (*see*, Figure 2A). In another example, a 0.9 kb restriction segment between restriction sites *HincII* and *HpaII* of plasmid pFN2 shown in Figure 3 hybridizes to the *repA* probe of plasmid pFN1, indicating that *repA* nucleic acid of plasmid pFN2 is located within this 0.9 kb restriction segment.

Any other plasmids derived from, *e.g.*, any one of pFN1, pFN2 and pFN3 are within the embodiments of the invention, as long as they contain necessary regions for

replication in *Fusobacterium*. For example, these derivatives can be derived either by the deletion of unnecessary regions of replication from these plasmids, or by insertion or addition of any other DNA to these plasmids, or substitution of sequences in the regions for replication. Therefore, embodiments of the invention are not limited to plasmids pFN1, pFN2 and pFN3 themselves, but include other derivative plasmids by modifying these plasmids as well as other recombinant plasmids obtained by insertion of other nucleic acids, *e.g.*, marker genes, promoters, other origin of replication sequences, *etc.* The recombinant methods for production of these plasmids and vectors are described in detail below.

V. Construction of Recombinant Plasmids and Vectors

In yet another aspect, the invention provides recombinant plasmids and vectors that are functional in *Fusobacterium*. The invention also provides plasmids and vectors that are functional in *Fusobacterium* as well as in other microorganisms, such as *E. coli*. The plasmids and vectors of the invention can be used for cloning and expressing *Fusobacterium* genes. The plasmids and vectors of the invention can also be used to express foreign genes in *Fusobacterium* or in other microorganisms.

The recombinant plasmids and vectors can be produced by joining the nucleic acids, plasmids or fragments described herein and other nucleotide sequences using recombinant methods known in the art. Typically, a plasmid or a vector of the invention comprises an origin of replication functional in *Fusobacterium* (*e.g.*, *F. nucleatum*), convenient restriction endonuclease sites and one or more selectable markers. Other elements can be included depending on the desired use of a vector.

A plasmid or a vector of the invention can comprise any origin of replication sequences described herein. For example, a plasmid or a vector comprises an origin of replication comprising at least two copies of the iteron, wherein the iteron has a nucleic acid sequence of SEQ ID NO:3. Preferably, a plasmid or a vector comprises an origin of replication comprising at least two to six copies of the iteron having a nucleic acid sequence of SEQ ID NO:3. More preferably, a plasmid or vector of the invention comprises a nucleic acid sequence of SEQ ID NO:4. Alternatively, restriction fragments of plasmids pFN1, pFN2 or pFN3 that contain the origin of replication sequences can be ligated to provide a recombinant plasmid or vector of the invention.

A plasmid or a vector of the invention can also comprise any suitable replication nucleic acids for *F. nucleatum* described herein that are compatible with the origin of

000001 000002 000003 000004 000005 000006 000007 000008 000009 000010 000011 000012 000013 000014 000015 000016 000017 000018 000019 000020

replication sequences of the plasmid or vector, provided that the repA nucleic acids have other than the nucleic acid sequence of SEQ ID NO:5 (a nucleic acid of plasmid pAD52 that encodes a RepA homolog; *see*, GenBank Accession No. AF022647, the nucleotide positions 1108 through 2337). For example, a plasmid or a vector of the invention comprises a repA nucleic acid that encodes an amino acid sequence that is at least about 80%, preferably at least about 85%, more preferably at least about 90%, most preferably at least about 95%, 96%, 97%, 98% or 99%, identical to an amino acid as set forth in SEQ ID NO:1 over a region of at least about 50 amino acids in length. Alternatively, a plasmid or a vector of the invention comprises a repA nucleic acid that is at least about 80%, preferably at least about 85%, more preferably at least about 90%, most preferably at least about 95%, 96%, 97%, 98% or 99% identical to a nucleic acid as set forth in SEQ ID NO:2. In preferred embodiments, the region of identity extends over a region longer than 50 amino acids or nucleotides, preferably over a region of at least about 75 amino acids or nucleotides, more preferably over a region of at least about 100 amino acids or nucleotides, most preferably over the full length of the repA amino acid or nucleotide sequence. In preferred embodiments, a plasmid or a vector of the invention comprises a repA nucleic acid that encodes a polypeptide having the amino acid sequence of SEQ ID NO:1 or has a nucleic acid sequence of SEQ ID NO:2. These repA nucleic acids are operably linked to a promoter and other regulatory sequences in a plasmid or a vector in any suitable manner.

In yet another embodiment, a vector can comprise any combination of origin of replication sequences and repA nucleic acid sequences described above. For example, a vector can comprise an origin of replication comprising a nucleic acid sequence of SEQ ID NO:4 and a repA nucleic acid sequence comprising a nucleic acid sequence of SEQ ID NO:2. In another example, a vector can comprise a 2.36 kb fragment between restriction sites *AvrII* and *ScaII* derived from plasmid pFN1 (or conservatively modified variants thereof) and a repA nucleic acid comprising a nucleic acid sequence of SEQ ID NO:2. In yet another example, a vector can comprise a nucleic acid comprising an origin of replication at nucleotide position 3936 to 4481 of plasmid pFN1 and a 0.9 kb fragment between restriction sites *HincII* and *HpaII* derived from plasmid pFN2 (or conservatively modified variants thereof) that comprises repA nucleic acid sequences.

In some embodiments, a plasmid or a vector can be a shuttle vector comprising additional origin of replication sequences so that it can replicate in

Fusobacterium as well as in another microorganism. Such sequences are well known for a variety of microorganisms, including, *e.g.*, Gram-negative or Gram-positive bacteria. For instance, the origin of replication sequences in pBR or pUC series plasmids can be ligated to produce a shuttle vector of the invention. Preferably, additional origin of
5 replication sequences are selected so that a shuttle vector can replicate in *E. coli*. Such shuttle vectors would have high applicability, since *E. coli* is known to be an efficient host for DNA amplification and manipulation.

The plasmids or vectors of the invention can also comprise selective marker genes to allow selection of host cells that have been transformed with a plasmid or
10 a vector. These marker genes encode a protein necessary for the survival or growth of transformed host cells grown in selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that confer resistance to antibiotics or other toxins, such as erythromycin, clindamycin, ampicillin, neomycin, kanamycin, penicillin,
15 cefotaxime, imipenem, metronidazole, streptomycin, chloramphenicol, or tetracycline. Marker genes that are particularly useful in selection of transformed *F. nucleatum* include, *e.g.*, a gene that encodes resistance to clindamycin. Alternatively, selective markers may encode proteins that complement auxotrophic deficiencies or supply critical nutrients not available from complex media. A number of selective markers are known to
20 those skilled in the art and are described for instance in Sambrook *et al.*, *supra*.

In some embodiments, one or more transcription cassettes comprising a nucleic acid of interest and control sequences can be included to provide an expression vector of the invention. Commonly used control sequences include promoters for transcription initiation, optionally with an operator, along with ribosome binding site
25 sequences. These control sequences are operably linked to a nucleic acid of interest to enable transcription and translation of a polypeptide of interest.

As a promoter, either constitutive or regulated promoters can be used in the present invention, and the selection of a promoter depends on the host cell selected for expression. For example, expression of a nucleic acid of interest in *E. coli* would require
30 a promoter that is functional in *E. coli*. A number of suitable promoters for *E. coli* are known in the art (*see, e.g.*, Sambrook *et al.*, *supra*). If expression of a nucleic acid in *F. nucleatum* is desired, a promoter that is functional in *F. nucleatum* is included in a plasmid or a vector of the invention. Such promoters can be obtained from genes that have been cloned from *F. nucleatum*, or heterologous promoters functional in *F.*

nucleatum may be used. Exemplary genes that have been cloned from *F. nucleatum* include, e.g., a *fomA* gene (see, Haake & Wang, *Archs. Oral Biol.* 42:19-24 (1997)), a *fipA* gene (see, Demuth *et al.*, *Infec. Immun.* 64:1335-1341 (1996)), *etc.* The promoters for these genes can be used to express a nucleic acid of interest in *F. nucleatum*.

5 A ribosome binding site (RBS) can also be included in the transcription cassettes of the invention. An RBS in *E. coli*, for example, consists of 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon (Shine & Dalgarno, *Nature* 254:34 (1975); Steitz, *In Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger), vol. 1, p.349, Plenum Publishing, NY (1979)).

10 A nucleic acid of interest in a transcription cassette can include either *F. nucleatum* nucleic acids or foreign nucleic acids. For example, *F. nucleatum* nucleic acids that encode immunosuppressant proteins, such as *fomA*, *fipA*, or subsequences thereof, can be included in a transcription cassette. Alternatively, a number of variety of foreign genes can be included in a transcription cassette. For example, leukotoxin genes, such as *ltxA*, from *Actinobacillus actinomycetemcomitans*, *etc.*, or subsequences thereof can be included in a transcription cassette. In another example, protease genes from *Porphyromonas gingivalis*, such as *rgpA*, *rgpB*, *kgp*, *prtT*, *etc.*, or subsequences thereof can be included in a transcription cassette. In yet another example, cholera toxin genes, such as *ctxA*, *ctxB*, or any subunits of cholera toxins, or subsequences thereof can be included in a transcription cassette. If a foreign gene is isolated from a pathogenic microorganism (e.g., a leukotoxin, endotoxin, a cholera toxin, *etc.*), *F. nucleatum* transformed with such foreign genes or antigenic fragments thereof can be used as a vaccine delivery system to stimulate mucosal immunity against those pathogenic microorganisms.

25 The nucleic acids of interest can be expressed intracellularly or can be secreted from the cell. Intracellular expression often results in high yields. If necessary, the amount of soluble polypeptides of interest can be increased by performing refolding procedures (see, e.g., Sambrook *et al.*, *supra*; Marston *et al.*, *Bio/Technology* 2:800 (1984); Schoner *et al.*, *Bio/Technology* 3:151 (1985)). In embodiments in which the polypeptides of interest are secreted from the cell, either into the periplasm or into the extracellular medium, the nucleic acid of interest is linked to another nucleic acid that encodes a cleavable signal peptide sequence. The signal sequence directs translocation of the polypeptide of interest through the cell membrane.

30 The polypeptides of interest can also be produced as fusion proteins to aid purification of the polypeptides. For example, the DNA encoding the polypeptide of

interest may be fused with a nucleotide sequence that contains an affinity tag so that purification of recombinant polypeptides can be simplified. For example, multiple histidine residues encoded by the tag allow the use of metal chelate affinity chromatography methods for the purification of fusion polypeptides. Other examples of affinity tag molecules include, Strep-tag, Pinpoint, maltose binding protein, glutathione S-transferase, *etc.* See, *e.g.*, Glick and Pasternak, *Molecular Biotechnology Principles and Applications of Recombinant DNA*, 2nd Ed., American Society for Microbiology, Washington, DC (1999).

Construction of suitable plasmids or vectors containing one or more of the above listed components employs standard ligation techniques as described in the reference cited above. Isolated plasmids or DNA fragments are cleaved, tailored, and re-ligated in the form desired to generate the plasmids or vectors desired. To confirm correct sequences in a plasmid or a vector constructed, the plasmid or the vector can be analyzed by standard techniques such as by restriction endonuclease digestion, and/or sequencing according to known methods.

VI. Host Cells, Transformation of Host Cells and Protein Purification

A number of host cells can be used for transformation with the vectors of the invention. Examples of useful host cells include *Fusobacterium*, *Eschericia*, *Leptotrichia*, *Streptococcus*, *Staphylococcus*, *Clostridium*, *etc.* Suitable *Fusobacterium* hosts include *F. nucleatum*, *F. necrophorum*, *F. varium*, *F. periodonticum*, *etc.* In particular, *F. nucleatum* is a preferred host cell. Suitable *F. nucleatum* strains include, *e.g.*, *F. nucleatum* strains 12230 or 10113 (both of which are a gift from S. Finegold, West Los Angeles, VA), *F. nucleatum* having ATCC Accession Nos. 10953 or 23726, *etc.* Suitable *Eschericia* hosts include the following strains: JM101, RR1, DH5 α , and others. These examples are illustrative rather than limiting.

The host bacterial cells can be transformed with the vectors of the present invention using standard methods appropriate to such cells. These methods include, *e.g.*, electroporation, calcium chloride methods, polyethylene glycol methods, *etc.* Cells transformed by the vectors can be selected by, *e.g.*, resistance to antibiotics conferred by genes contained on the vector.

For host cells that are transformed with an expression vector, the polypeptides that are expressed by the host cells can be purified according to standard

procedures of the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis and the like (*see, generally*, R. Scopes, Protein Purification, Springer-Verlag, N.Y. (1982), Deutscher, Methods in Enzymology Vol. 182: Guide to Protein Purification, Academic Press, Inc. N.Y. (1990)).

5

EXAMPLES

The following examples are provided by way of illustration only and not by way of limitation. Those of skill in the art will readily recognize a variety of noncritical parameters that could be changed or modified to yield essentially similar results.

10

The relevant characteristics of the bacterial strains and plasmids used in the examples below are summarized in Table 4.

TABLE 4

Bacterial strain Or plasmid	Relevant characteristics ¹	Source/Reference
<i>E. coli</i> DH5α	Ery ^s	Life Technologies, Gaithersburg, MD
<i>F. nucleatum</i> 12230	Cln ^s , transtracheal isolate, source of pFN1	S. Finegold, Wadsworth Anaerobe Lab, West Los Angeles VA Medical Center, Los Angeles, CA
<i>F. nucleatum</i> 10113	Clns, clinical isolate, source of pFN2	S. Finegold, Wadsworth Anaerobe Lab, West Los Angeles VA Medical Center, Los Angeles, CA
<i>F. nucleatum</i> ATCC 10953	Cln ^s , gingival isolate, source of pFN3	American Type Culture Collections, Rockville, MD
<i>F. nucleatum</i> ATCC 23726		American Type Culture Collections, Rockville, MD
pBluescript SK(-)	Amp ^r , 3.0 kb <i>E. coli</i> vector	Stratagene Cloning Systems
pVA2198	Ery ^r , 9.2 kb plasmid with the <i>ermF-ermAM</i> cassette	F. Macrina (<i>see, Fletcher et al., Infect. Immun.</i> 63:1521-1528 (1995))
pFN1	Cln ^s , 5.9 kb native <i>F. nucleatum</i> plasmid	This study
pFN2	Cln ^s , 7.2 kb native <i>F. nucleatum</i> plasmic	This study
pFN3	Cln ^s , 11.1 kb native <i>F. nucleatum</i>	This study

¹ Abbreviations: amp, ampicillin; ery, erythromycin; cln, clindamycin.

	plasmid	
pHS17	Amp ^r , Ery ^r , 10.0 kb plasmid consisting pFN1, <i>ermF-ermAM</i> , and pBluescript minus its ampicillin resistance determinant	This study
pHS19	Amp ^s , Ery ^r , 4.1 kb plasmid consisting of <i>ermF-ermAM</i> and pBluescript minus its ampicillin resistance determinant	This study

Example 1. Isolation and Characterization of *F. nucleatum* plasmids

Three native plasmids pFN1, pFN2, and pFN3 (*see, e.g.,* Table 1; Fig. 1A) were isolated from strains of *F. nucleatum* using routine techniques (Wizard Plus Minipreps; Promega, Madison, WI; Qiagen Midi Preps, Qiagen Inc., Valencia, CA) and visualized on ethidium-stained 0.8% agarose gels. Restriction endonuclease mapping was accomplished using standard recombinant DNA technologies (*see, Kinder Haake et al., Arch. Oral Biol.* 42:19-24 (1997)). The results of restriction endonuclease mapping demonstrated that the plasmids varied in size and in the occurrence of several restriction endonuclease sites (*see, Fig. 1A*), suggesting that the plasmids were unrelated.

Southern hybridization studies were performed under conditions described in Kinder *et al.*, *Gene* 136:271-275 (1993), and the results indicated that pFN1 and pFN2 share homology with each other, but not with pFN3 (*see, Fig. 1B-C*). Nitrocellulose blots of plasmid and chromosomal DNA preparations from the plasmid-containing host strains were probed with pFN1 and pFN3 DNA. The pFN1-probe hybridized to pFN1 and pFN2, but not pFN3 DNA (*see, Fig. 1B*), whereas the pFN3-probe hybridized only to pFN3 (*see, Fig. 1C*). No hybridization to chromosomal DNA from any of the host strains was evident (data not shown).

The strain harboring pFN3, ATCC 10953, was previously reported to lack plasmid DNA (*see, McKay et al., Plasmid* 33:15-20 (1995)). Due to this discrepancy we obtained a new culture from ATCC and confirmed the presence of pFN3 in this strain. These data reveal the existence of two non-homologous groups of plasmids indigenous to *F. nucleatum*, the first represented by pFN1 and pFN2, and the second represented by pFN3.

Example 2. Determination and Analysis of pFN1 DNA Sequence and Partial pFN3 DNA Sequence

Due to its small size and superior plasmid yields, pFN1 was chosen for further analysis. The DNA sequence (GenBank Accession No. AF159249) was determined for both strands. Analysis of the compiled sequence revealed a circular structure of 5887 bp with 23% G+C content and seven putative open reading frames (ORFs defined as ≥ 150 bp; Fig. 2A). Similarity searches were performed using the National Center for Biotechnology Information BLAST server (Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990); Altschul *et al.*, *Nuc. Acids Res.* 25:3389-3402 (1997)). The sequence of pFN1 was highly homologous to the sequence of a 6281 bp *F. nucleatum* plasmid (pAD52, GenBank Accession No. AF022647). No similarity was found to any gene encoding antibiotic resistance or other selectable phenotypic marker. Antibiotic susceptibility testing indicated that the pFN1 host strain *F. nucleatum* 12230 was susceptible to penicillin G, tetracycline, chloramphenicol, clindamycin, cefotaxime, ampicillin/sulbactam, imipenem, metronidazole and streptomycin, and resistant to erythromycin at 25 μ g/ml as is common in *F. nucleatum* (see, Brazier *et al.*, *J. Appl. Bacteriol.* 71:343-346 (1991)). These data suggested that pFN1 is a cryptic plasmid with respect to antibiotic resistance, comparable to previous findings with this group of plasmids (see, McKay *et al.*, *Plasmid* 33:15-20 (1995)).

ORF1 is related to DNA relaxase (mobilization) proteins which mediate the initiation of conjugal transfer of plasmid DNA (Ilyina *et al.*, *Nuc. Acids Res.* 20:3279-3285 (1992)). Alignment of the complete predicted amino acid sequences using Clustal W (Higgins, *Methods Mol. Biol.* 25:307-318 (1994); Thompson *et al.*, *Nuc. Acids Res.* 22:4673-4680 (1994)) of ORF1 with *Staphylococcus* plasmid relaxases demonstrated 23 to 29% identity and 30 to 34% similarity. Homology to the four regions of the consensus sequence defined for relaxase proteins (Ilyina *et al.*, *supra*) was evident (data not shown).

ORF5 analyses indicated that it is related to plasmid replication proteins, including *Lactobacillus acidophilus* plasmid pLA103 (Kanatani *et al.*, *FEMS Microbiol. Lett.* 133:127-130 (1995)), *Staphylococcus aureus* plasmid pJEl (Berg *et al.*, *J. Bacteriol.* 180:4350-4359 (1998)), and *Pediococcus halophilus* pUCL287 (Benachour *et al.*, *FEMS Microbiol. Lett.* 128:167-176 (1995)). Alignment of the complete ORFs of homologues with pFN1 ORF5 demonstrated 10 to 19% identity and 21 to 34% similarity. The association of ORF5 with replication was strongly supported by analyses of the upstream DNA sequence, which demonstrated six perfect 22 bp direct repeats ("iterons") preceded

by an approximately 200 bp A-T rich region (Fig. 2B). Multiple putative DnaA binding sites were also identified, based on matching eight of the nine bp comprising the DnaA binding consensus sequence (Schaefer *et al.*, *Mol. Gen. Genet.* 226:34-40 (1991)). This organization is characteristic of the origin of replication of iteron-regulated theta-replicating plasmids (Helinski *et al.*, Replication control and other stable maintenance mechanisms of plasmids, p. 2295-2324. In Neidhardt (ed.), *Escherichia coli* and *Salmonella*, 2nd ed., vol. 2. ASM Press, Washington, D. C. (1996)). A general model of replication initiation involves the binding of the plasmid replication protein to the iteron sequences resulting in structural changes, including melting of the adjacent A-T rich region, to form an open complex. The replication protein, possibly in conjunction with the host DnaA protein, is then responsible for guiding host replication proteins into the open complex (Helinski, *et al.*, *supra*; Neidhardt, *supra*). It is also significant that the pFN1 replication protein homologue was related to the replication protein of pUCL287, which has been shown to utilize a theta mode of replication (Benachour *et al.*, *FEMS Microbiol. Lett.* 128:167-176 (1995)).

A partial DNA sequencing was also performed with plasmid pFN3. The partial nucleotide sequence of plasmid pFN3 is as shown in SEQ ID NO:14. The sequence is degenerate in that there were ambiguous bases (designated as "N") within the sequence. The sequence analysis was performed using a clone of a fragment of plasmid pFN3 (clone designated as pSY1).

Example 3. Homology Studies Between Plasmids pFN1 and pFN2

To further characterize the regions of homology between pFN2 and pFN1, a Southern blot analysis was done using probes specific for two pFN1 genes. Plasmids pFN1 and pFN2 were subject to restriction endonuclease digestion using different restriction endonucleases and then electrophoresis on .8% agarose gel. The DNA was transferred to nitrocellulose, hybridized with the radiolabeled probe, and washed according to standard techniques (*see*, Kinder *et al.*, *Gene* 136:271-5 (1993)). Partial restriction maps are as shown in Figure 3. As evident in Figure 3, different restriction endonucleases were used for these homology studies.

The repA and rlx gene sequences were amplified from pFN1 DNA using the polymerase chain reaction with oligonucleotide primers specific for these regions. The DNA probes used were amplified by polymerase chain reaction from pFN1 using

custom primers (rlx primers: 5'-CCTGG TGAAGTAGATGAAG-3' (SEQ ID NO:7), 5'-TTAGTTTTAGCAATGGAAG-3' (SEQ ID NO:8), repA primers: 5'-ATGCTGGAGTGTGATATG-3' (SEQ ID NO:9), 5'-GTTGATTTTCCACTTTCGG-3' (SEQ ID NO:10); Gibco Life Technologies, Grand Island, NY).

5 Hybridization of the repA and rlx probes with pFN1 digests gave the expected results (Fig. 4, lanes 1 and 2; Fig. 5). Both probes hybridized with the linear 5.9 kb pFN1 DNA band (Fig. 4, lanes 1). The repA probe hybridized to the 3.1 kb band of the pFN1 StyI digested DNA (Fig. 4, lane 2, left and middle panels) whereas the rlx probe hybridized to the 2.7 kb band (Fig. 4, lane 2, left and right panel).

10 Hybridization of both the pFN1 repA and rlx gene probes were evident with the linear pFN2 band at 7.2 kb (Fig. 4, lane 3, middle and right panels). The pFN1 repA probe (Fig. 4, middle panel) hybridized with a 2.8 kb band in the *HincII* digest (lane 4), a 0.9 kb band in the *HincII/HpaII* digest (lane 5), and a 5.5 kb band in the *AvrII* digest (lane 6). This pattern of hybridization indicates that the region of homology of the pFN1 repA gene on pFN2 is localized to the 0.9 kb *HincII/HpaII* fragment (Fig. 5).

15 The pFN1 rlx gene probe (Fig. 4, right panel) demonstrated hybridization with the 3.4 kb bands in the *HincII* (lane 4) and *HincII/HpaII* (lane 5) digests, and with a 5.5 kb band in the *AvrII* digest (lane 6). These data indicate that the region with homology to the pFN1 rlx gene is localized to the 2.2 kb *HincII-AvrII* fragments of pFN2 (Fig. 5).

Example 4. Isolation of repA Nucleic Acids and Origin of Replication Sequences

25 The repA nucleic acids and origin of replication sequences of the invention can be obtained using a number of methods known in the art. This example illustrates the isolation of repA nucleic acids and origin of replication sequences using PCR methods. PCR reactions are performed as described by the manufacturer (*e.g.*, Boehringer Mannheim, Montreal). The primers used to amplify the repA and origin of replication sequences were based on the sequence information of plasmid pFN1 (*see*, SEQ ID NO:6).

30 To amplify repA gene sequence (pFN1 nucleotide positions 4429 to 158 [note that this goes through the "0" position]), the following primers can be used.

Forward: 5'-GAC ATT AAG TGA AAA AG-3' (SEQ ID NO:16)

Reverse: 5'-ATG CTG GAG TGT GAT ATG-3' (SEQ ID NO:17)

To amplify the origin of replication, including the AT rich region, the

iteron repeat sequences and the putative DnaA binding sites (positions 3677 to 4487 of pFN1), the following primer can be used.

Forward: 5'-ACG GAT ACT TTG TTG CT-3' (SEQ ID NO:18)

Reverse: 5'-TAT CCT TTA CAT TTA-3' (SEQ ID NO:19)

To amplify the origin of replication and the repA gene, combined (pFN1 nucleotide positions 3677 to 158 [note that this sequence goes through the "0" position]), the following primer can be used.

Forward: 5'-ACG GAT ACT TTG TTG CT-3' (SEQ ID NOS:20)

Reverse: 5'-ATG CTG GAG TGT GAT ATG-3' (SEQ ID NO:21)

The PCR products are purified on a spin column (e.g., S-300 spin column; Pharmacia Biotech) and are sequenced to confirm isolation of the correct sequences.

Example 5. Transformation of *F. nucleatum* with the Shuttle Plasmid pHS17

The shuttle plasmid pHS17 (Fig. 1A) was constructed sequentially in *E. coli* DH5 α cells (Life Technologies, Gaithersburg, MD) as follows: *AvrII*-digested pFN1 was cloned into the *XbaI* site of pBluescript; an *ermF-ermAM* cassette (Fletcher *et al.*, *Infect. Immun.* 63:1521-1528 (1995); *see also*, GenBank Acc. No. AF219231) was added by cloning into *KpnI/PstI* sites; and the pBluescript ampicillin resistance determinant was deleted by digestion with flanking *BspHI* sites. The resulting construct included both *E. coli* and *F. nucleatum* origins of replication, from pBluescript and pFN1, respectively. The junctions of DNA fragments joined by cloning were confirmed by DNA sequencing, and the phenotypic properties of the construct were confirmed on selective media.

Transformation studies were performed with plasmid DNA isolated by alkaline lysis/column purification techniques (Wizard Plus Minipreps, Promega, Madison, WI; Qiagen Midi Preps, Qiagen Inc. Valencia, CA) and further purified by cesium chloride ethidium bromide density gradient centrifugation (*see*, Sambrook, *supra*). Bacterial cells were washed, resuspended in electroporation buffer according to methods of Fletcher *et al.*, *supra*, at a calculated optical density of 0.60. 100 μ l aliquots were electroporated using standard techniques (*see*, Sreenivasan *et al.*, *Infect. Immun.* 59:4621-4627 (1991)). The electroporated cells were immediately diluted with 0.9 ml of Columbia broth (BBL Microbiology Systems, Cockeysville, MD) with MgCl₂, and the number of viable cells were determined by plating a diluted aliquot on non-selective media. The transformation mix was incubated anaerobically followed by plating on

Columbia agar (BBL Microbiology Systems) with clindamycin. Variables examined included the bacterial cell growth phase (early log, mid-log, and stationary phase), the source of pHS17 DNA (heterologous versus homologous host sources), electroporation parameters (resistance of 50 to 500 Ω ; field strength of 24 or 25 kV/cm; capacitance of 25 or 50 μ F), the concentration of $MgCl_2$ in the Columbia broth (0.5, 1.0, 2.0 mM), and the clindamycin concentration used in the selective media (0.2 or 0.4 μ g/ml).

Transformation of *F. nucleatum* ATCC 10953 with pHS17 were successful using previously defined conditions (Fletcher, *Appl. Environ. Microbiol.* 52:672-676 (1986); Rosey *et al.*, *J. Bacteriol.* 177: 5959-5970 (1995); Sreenivasan *et al.*, *supra*).

Preliminary results indicated optimal recovery of transformant with a field strength of 25 kV/cm, a capacitance of 25 μ F, and resistance ranging from 200 to 400 Ω . Analysis of the transformants revealed the presence of pHS17, and the ATCC 10953 native plasmid pFN3. The two plasmids were easily distinguished based on their sizes and restriction endonuclease digestion patterns (Fig. 6). Electroporation controls included non-electroporated cells with or without the addition of DNA as well as electroporated cells without DNA added, and all yielded negative results. Electroporation with pHS19 also yielded negative results, suggesting that pFN1 is essential for replication in *F. nucleatum*.

Transformation efficiency was dependent on the pHS17 DNA source. The transformation efficiency using 1 μ g of plasmid DNA ranged from 1.6 to 2 x 10² transformants per μ g of DNA from the homologous *F. nucleatum* host, as compared to no transformants with DNA from the heterologous *E. coli* host as shown in Table 4 below.

TABLE 4

	Resistance ^b (Ohms):	Heterologous plasmid DNA ^a : pHS17		Homologous plasmid DNA ^a : pHS17 + pFN3
		1 μ g DNA	5 μ g DNA	1 μ g DNA
Transformation Efficiency ^c	400	0	1.2 x 10 ¹	1.6 x 10 ²
	300	0	0	1.9 x 10 ²
	200	0	1	2.0 x 10 ²

^a Heterologous plasmid DNA was isolated from *E. coli* strain KTK5 (pHS17). Homologous plasmid DNA was isolated from *F. nucleatum* strain KH21 (pHS17, pFN3). The quantitation of plasmid DNA is based on total DNA in the preparation.

^b See above for additional electroporation parameters. Outgrowth period was approximately 12 hours.

^c Calculated as the number of transformants per ml/number of μg of DNA (see, Sreenivasan *et al.*, *Infect. Immun.* 59(12):4621-4627).

The average cells remaining after electroporation at 400, 300 and 200 ohms was 0.39×10^8 , 1.1×10^8 and 2.1×10^8 , respectively.

5 Transformation efficiency was optimal at a resistance setting of 200 Ω using the homologous host DNA, although pronounced difference were not evident over the range examined. Transformation with *E. coli* pHS17 DNA at 5 μg was demonstrated; however, the efficiency was still less than that observed with 1 μg of homologous DNA. The 100-fold or greater increase in transformation efficiency with homologous DNA suggests the
10 presence of a functional restriction-modification system in *F. nucleatum* ATCC 10953. Restriction-modification systems in *F. nucleatum* have been previously reported (see, Leung *et al.*, *Nuc. Acids Res.* 6:17-25 (1979); Lui *et al.*, *Nuc. Acids Res.* 6:1-15 (1979)).

Growth phase also influenced the transformation efficiency, but to a lesser extent than the DNA source. Increased transformation efficiencies were transformation
15 efficiencies were routinely obtained with early log phase cells. For example, in one experiment using early log, mid-log and stationary phase recipient cells and an outgrowth period of 5 hours (approximately 2 generations), the transformation efficiencies were 7.2, 4.8 and 5.0×10^3 , respectively. No significant differences were observed with variations in the concentration of MgCl_2 in the outgrowth broth, or with 0.2 versus 0.4 $\mu\text{g}/\text{ml}$
20 clindamycin in the selective agar media.

In addition, *F. nucleatum* subspecies *nucleatum* ATCC 23726 yielded transformants with an efficiency ranging from 1.2×10^4 transformants per ml per μg of DNA. Analyses revealed the presence of plasmid DNA from the transformants that was consistent with pHS17 in size and restriction endonuclease digestion pattern.

25 **Example 6. Stability of Shuttle Plasmid in *F. nucleatum* Transformants**

The structural stability of pHS17 in representative transformants was evaluated by restriction endonuclease mapping, PCR-amplification of pHS17-specific DNA regions, and Southern analysis of the transformant DNA with pFN1 and pHS17
30 DNA probes (data not shown). In all of the analyses done, no evidence of DNA rearrangement or deletion was detected. The segregational stability of pHS17 was examined in the transformant strain KH21 maintained in liquid cultures without antibiotic selection as described in Roberts *et al.*, *J. Bacteriol.* 174:8119-8132 (1992). After 100 generations the percentage loss of plasmid per generation was 0.02, with an average of

98% of the viable cells demonstrating the clindamycin-resistance phenotype. The shuttle plasmid was present in all colonies subcultured at baseline and after 100 generations, with no evidence of DNA rearrangement or deletion. Thus, pHS17 was found to be both structurally and segregationally stable in the *F. nucleatum* host cell background.

- 5 Interestingly, both pHS17 and pFN3 were stably maintained in the transformants, indicating that these two plasmids are compatible, and that pFN3 may be useful in developing plasmid vectors for use in conjunction with pFN1-derived plasmids.

The present invention provides novel materials and methods related to
10 *Fusobacterium*. While specific examples have been provided, the above description is illustrative and not restrictive. Any one or more of the features of the previously described embodiments can be combined in any manner with one or more features of any other embodiments in the present invention. Furthermore, many variations of the
15 invention will become apparent to those skilled in the art upon review of the specification. The scope of the invention should, therefore, be determined not with reference to the above description, but instead should be determined with reference to the appended claims along with their full scope of equivalents.

All publications and patent documents cited in this application are
incorporated by reference in their entirety for all purposes to the same extent as if each
20 individual publication or patent document were so individually denoted. By their citation of various references in this document Applicants do not admit any particular reference is "prior art" to their invention.